Duality and the Modular Group in the Quantum Hall Effect

Brian P. Dolan
Department of Mathematical Physics, National University of Ireland, Maynooth, Ireland
and
Dublin Institute for Advanced Studies, 10 Burlington Rd., Dublin, Ireland
e-mail: bdolan@thphys.may.ie
(Revised version: 25th December 1998)

We explore the consequences of introducing a complex conductivity into the quantum Hall effect. This leads naturally to an action of the modular group on the upper-half complex conductivity plane. Assuming that the action of a certain subgroup, compatible with the law of corresponding states, commutes with the renormalisation group flow, we derive many properties of both the integer and fractional quantum Hall effects including: universality; the selection rule $|p_1q_2 - p_2q_1| = 1$ for transitions between quantum Hall states characterised by filling factors $\nu_1 = p_1/q_1$ and $\nu_2 = p_2/q_2$; critical values of the conductivity tensor; and Farey sequences of transitions. Extra assumptions about the form of the renormalisation group flow lead to the semi-circle rule for transitions between Hall plateaux.

PACS nos: 73.40.Hm, 05.30.Fk, 02.20.-a

The purpose of this letter is to explore the consequences of the proposal, made in [1] and examined further in [2] [3], that the hierarchical structure of the zero temperature integer and fractional quantum Hall effects can be interpreted in terms of the properties of a subgroup of the modular group, $Sl(2,\mathbf{Z}) := \Gamma(1)$ — specifically the subgroup which consists of elements of $\Gamma(1)$ whose bottom left entry is even, sometimes denotd $\Gamma_0(2)$ in the mathematical literature. This group acts on the upper-half complex plane, parameterised by the complex conductivity, $\sigma = \sigma_{xy} + i\sigma_{xx}$, in units of $\frac{e^2}{h}$, and is generated by two operations, $T : \sigma \to \sigma + 1$ and $X : \sigma \to \frac{\sigma}{2\sigma+1}$. If $\gamma = \begin{pmatrix} a & b \\ 2c & d \end{pmatrix} \in \Gamma_0(2)$, with a, b, c, and $d \in \mathbf{Z}$ and $ad - 2cb = 1$, then $\gamma(\sigma) = \frac{a\sigma+b}{2c\sigma+d}$. Thus $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $X = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. Some consequences of this assumption for the phase diagram in the σ-plane were examined in [2] and in the second of these references the author notes that there is a connection with the work of Kivelson, Lee and Zhang [4], but remarks that the comparison between [2] and [4] is not immediate. One of the aims of this paper is to explore the relation between these two approaches.

Following [1]— [3], it will be assumed that the phase diagram for the quantum Hall effect can be generated by the action of $\Gamma_0(2)$ on the upper-half σ plane. This immediately implies the 'law of corresponding states' of [4] and [5]. At Hall plateaux we have $\sigma_{xx} = 0$ and $\sigma_{xy} = s$ where s is a ratio of two mutually prime integers, with odd denominator (note that s is being used here to label the quantum phases and is denoted by s_{xy} in [4]). Plateaux can be related to each other by repeated action of T and X. At the center of the plateaux, the filling factor, ν, is equal to the ratio $s = p/q$ and $T : \nu \to \nu + 1$ is the Landau level addition transformation of [4] while $X : \nu \to \frac{\nu}{2\nu+1}$ is the flux attachment transformation. The particle-hole transformation $\nu \to 1 - \nu$, can be realised as the outer auto-morphism $\sigma \to 1 - \bar{\sigma}$ acting on the upper-half plane, where $\bar{\sigma} = \sigma_{xy} - i\sigma_{xx}$ (it will be assumed throughout that the electron spins are split, for the spin degenerate case one must re-scale $\sigma \to 2\sigma$).

The upper-half σ-plane can be completely covered by copies of a single 'tile', or fundamental region (see e.g. [6]), related to each other by elements of $\Gamma_0(2)$. The fundamental region has cusps at 0 and 1, linked by a semi-circle of unit diameter, and consists of a vertical strip of unit width constructed above this semi-circle. By assumption all allowed quantum Hall transitions are images of the transition $\nu = 0 \to \nu = 1$ under some $\gamma \in \Gamma_0(2)$, and hence also linked by a semi-circle.

Each such semi-circle has a special point, in addition to the end points, which is a fixed point of $\Gamma_0(2)$ in the following sense. The point $\sigma^* = \frac{1+i}{2}$ is left fixed by $\gamma^* = \begin{pmatrix} 1 & -1 \\ 2 & -1 \end{pmatrix}$. Similarly the points obtained from σ^* by the other elements of $\Gamma_0(2)$, $\sigma^*_\gamma := \gamma(\sigma^*)$, are left fixed by $\gamma\gamma^*\gamma^{-1}$. It can be shown that the imaginary part, $\Im(\sigma^*_\gamma) \leq \frac{1}{2}$ or $\Im(\sigma^*_\gamma) = \infty$, $\forall\gamma$. The points σ^*_γ can be interpreted as critical points for the transition $\gamma(0) \leftrightarrow \gamma(1)$ if we further assume that the action of $\Gamma_0(2)$ commutes with the renormalisation group (RG) flow. For if σ^*_γ were not a RG fixed point, we could move to an infinitesimally close point $\phi(\sigma^*_\gamma) \neq \sigma^*_\gamma$ with a RG transformation, ϕ. Demanding $\gamma \circ \phi(\sigma^*_\gamma) = \phi \circ \gamma(\sigma^*_\gamma) = \phi(\sigma^*_\gamma)$ then implies that $\phi(\sigma^*_\gamma)$ is also left invariant by γ. But the fixed points of $\Gamma_0(2)$ are isolated, so there is no other fixed point infinitesimally close to σ^*_γ. Hence $\phi(\sigma^*_\gamma) = \sigma^*_\gamma$ and σ^*_γ must be a RG fixed point. The end points of the arches, at $\sigma = \nu$ with $\nu = p/q$ rational, are also fixed points of $\Gamma_0(2)$. For q odd these are stable Hall states. Note, how-

ever that a fixed point of the RG need not necessarily be a fixed point of $\Gamma_0(2)$ — but there is no experimental evidence of such extraneous fixed points of the RG.

Thus the fixed points of $\Gamma_0(2)$ must be fixed points of the RG, i.e. critical points. This leads to the topology of the flow diagram of [2], reproduced here in figures 1 and 2 where solid lines represent phase boundaries and dashed lines represent quantum Hall transitions. This implies the flow diagram proposed in [7], with its experimental support [8] and is also compatible with the phase diagram derived in [4]. That $\sigma^* = \frac{1+i}{2}$ is a critical point for the lowest Landau level was argued in [9]. Phase boundaries and transitions are represented by semi-circles in the figures, but this is not forced on us by the $\Gamma_0(2)$ symmetry. They could be distorted from this geometry, provided that all phase boundaries are copies of a distorted 'fundamental' phase boundary (running from $\frac{1}{2}$ to $\frac{1}{2}+i\infty$) under the action of $\Gamma_0(2)$. Similarly the dashed transition trajectories must all be copies of a distortion of the 'fundamental' arch spanning 0 and 1. Note, however that the *fixed points are immovable.* A useful aspect of the semi-circular arches used in the figures is that the intersection of any solid phase boundary with a dashed transition is a fixed point of $\Gamma_0(2)$, as are the end points of the arches (which are rational numbers or points at $\sigma = r + i\infty$ for integral or half-integral r). Any distortion from semi-circular geometry must leave the end points and intersections of phase boundaries and transition trajectories pinned at the fixed points of $\Gamma_0(2)$.

As in [4], the phase diagram generated by $\Gamma_0(2)$ determines which transitions are allowed and which are not. Thus, for example, $s : \frac{1}{3} \to 0$ is allowed while $s : \frac{1}{3} \to \frac{1}{7}$ is not. All allowed transitions are generated by acting on the arch passing through $\sigma = 0$ and $\sigma = 1$ by some, $\gamma \in \Gamma_0(2)$. This allows the derivation a selection rule for a transition $s_1 = p_1/q_1 \to s_2 = p_2/q_2$, where q_1 and q_2 are odd, and the pairs p_i and q_i $(i = 1, 2)$ are relatively prime (for brevity we shall not always distinguish below between s, labelling the quantum Hall phase, and ν, the filling factor, except where necessary for comparison with [4] — on the real axis, when $\sigma_{xx} = 0$, they are the same). We shall see that a transition is allowed if and only if $p_1q_2 - p_2q_1 = \pm 1$.

From the above assumptions we have (relabeling if necessary) $\nu_1 = \gamma(0)$, $\nu_2 = \gamma(1)$. Thus $\frac{p_1}{q_1} = \frac{b}{d}$ and $\frac{p_2}{q_2} = \frac{a+b}{2c+d}$ where $\gamma = \begin{pmatrix} a & b \\ 2c & d \end{pmatrix} \in \Gamma_0(2)$. Since $ad - 2cb = 1$, b and d are mutually prime, by an elementary result of number theory, hence (taking plus signs without loss of generality) $b = p_1, d = q_1$. Similarly $(a+b)d - (2c+d)b = 1$ implies that $a+b$ and $2c+d$ are mutually prime, hence $a+b = p_2$ and $2c+d = q_2$. Thus $\gamma = \begin{pmatrix} p_2 - p_1 & p_1 \\ q_2 - q_1 & q_1 \end{pmatrix}$ and the condition $det\gamma = 1$ then requires $p_2q_1 - p_1q_2 = 1$. The only possible exception to this rule would be a transition from $\nu = n \to \nu = m$ $(n, m \in \mathbf{Z})$, which could occur by going first from $\sigma = n$ to $\sigma = n + i\infty$ and then in to $\sigma = m$ from $\sigma = m + i\infty$. In a real experiment the maximum value of $|\sigma|$ would presumably be finite, due to impurities.

One can determine sequences of allowed transitions as follows. Suppose $\nu_0 = p_0/q_0$, with q_0 odd, is an allowed state, with p_0 and q_0 relatively prime. Consider the sequence $\nu_n = \frac{kn+p_0}{ln+q_0} := \frac{p_n}{q_n}$ for $n, k, l \in \mathbf{Z}$, where l is even (so that q_n is odd). Then $p_{n+1}q_n - p_nq_{n+1} = \pm 1 \Leftrightarrow kq_0 - lp_0 = \pm 1$. Thus a transition $\nu_{n+1} \to \nu_n$ is allowed provided $|kq_0 - lp_0| = 1$. In this way we can, for example, generate the three sequences

$$\frac{1}{3} \to \frac{2}{5} \to \frac{3}{7} \to \frac{4}{9} \to \frac{5}{11} \to \frac{6}{13} \to \dots$$

$$\dots \to \frac{7}{13} \to \frac{6}{11} \to \frac{5}{9} \to \frac{4}{7} \to \frac{3}{5} \to \frac{2}{3} \to 1$$

$$\frac{2}{3} \to \frac{5}{7} \to \frac{8}{11} \to \frac{11}{15} \to \dots \tag{1}$$

plus higher sequences obtained by adding an integer to each term in these sequences. Such sequences are called Farey sequences and their relevance to the quantum Hall effect was examined in [10]. Note that a given experiment may jump from one sequence to another. Thus

$$\dots \to \frac{3}{5} \to \frac{2}{3} \to \frac{5}{7} \to \dots$$

is observed in [11].

Each transition contains a critical point given by $\gamma(\sigma^*)$. Thus if $\gamma = \begin{pmatrix} a & b \\ 2c & d \end{pmatrix}$, the critical point is at

$$\sigma^*_\gamma = \frac{2ac + 2bc + ad + 2bd + i}{2d^2 + 4cd + 4c^2} = \frac{(p_1q_1 + p_2q_2) + i}{(q_1^2 + q_2^2)} \tag{2}$$

when the transition goes from $\nu_1 = \gamma(0) = b/d = p_1/q_1$ to $\nu_2 = \gamma(1) = \frac{a+b}{2c+d} = p_2/q_2$. The parameters of γ can be related to physical parameters as follows. Following [5], let η be the effective charge of the quasi-holes of a Hall state, $e^* = \eta$, θ the statistical parameter (i.e. the phase of the two quasi-particle wave function changes by $\pi\theta$ when the positions of the two particles are exchanged) and s be the Hall state, with magic filling factor $\nu = s$. Then the critical conductivity for a transition from $s = \nu$ to $s' = \nu - \eta^2/\theta$ is given by equation (26) of [5] (in dimensionless units)

$$\sigma_{xx} = \frac{\eta^2}{1+\theta^2}, \qquad \sigma_{xy} = s - \theta\frac{\eta^2}{1+\theta^2} \tag{3}$$

Equating these with the critical values in equation (2), there are two possibilities, depending on whether $\nu = \gamma(1)$ or $\gamma(0)$,

$$\text{i)} \qquad \nu = \frac{a+b}{2c+d} = \frac{p_2}{q_2} \quad , \quad \theta = \frac{d}{2c+d} = \frac{q_1}{q_2},$$

$$\eta^2 = \frac{1}{(2c+d)^2} = \frac{1}{q_2^2} \tag{4}$$

$$\text{ii)} \qquad \nu = \frac{b}{c} = \frac{p_1}{p_2} \quad , \ \theta = -\frac{(2c+d)}{d} = -\frac{q_2}{q_1},$$

$$\eta^2 = \frac{1}{d^2} = \frac{1}{q_1^2}. \tag{5}$$

In both cases we reproduce the result, that $\eta = \pm 1/q$, [12] and [13]. Note in passing that the transition from bosonic to fermionic conductivities given by equation (14) of reference [4] is implemented by the action of an element of $\Gamma(1)$ which is not in $\Gamma_0(2)$. Thus $\sigma = \gamma(\sigma^{(b)})$ where $\sigma^{(b)} = \sigma_{xy}^{(b)} + i\sigma_{xx}^{(b)}$ is the complex conductivity of the bosonic Chern-Simons action and $\gamma = \frac{1}{\eta}\begin{pmatrix} \eta^2 - \theta s & s \\ -\theta & 1 \end{pmatrix}$. The above discussion gives the explicit connection between the Chern-Simons analysis of [4] and the group theory analysis of [2].

We make a final comment about the 'semi-circle' law of reference [14] - [16]. By assumption, each quantum Hall transition can be obtained from the one between 0 and 1, passing through $\sigma^* = \frac{1+i}{2}$, by the action of some element of $\Gamma_0(2)$. Since $\Gamma_0(2)$ maps semi-circles built on the real axis into other such semi-circles we can deduce the 'semi-circle law' of reference [14] - [16] by making one extra assumption — that the 'fundamental' arch between 0 and 1 is a semi-circle. This implies that *all* other transitions are semi-circles and allows predictions to be made of the maximum values of σ_{xx} and ρ_{xx} in any allowed transition, $\nu_1 \to \nu_2$, as well as the values of σ_{xy} and ρ_{xy} at which they occur. Thus the maximum value of σ_{xx} is at $\sigma_{xx}^{max} = \frac{\nu_1 - \nu_2}{2}$, where $\sigma_{xy} = \frac{\nu_1 + \nu_2}{2}$ (where $\nu_1 > v_2$). In general, this does not coincide with the critical value $\sigma_\gamma^* = \gamma(\sigma^*)$, except for the integer transitions (table 1).

The maximum value of ρ_{xx} is found by constructing the semi-circle through $\frac{1}{\nu_1}$ and $\frac{1}{\nu_2}$ (provided neither vanishes). Thus $\rho_{xx}^{max} = \frac{1}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu_1}\right)$, where $\rho_{xy} = \frac{1}{2}\left(\frac{1}{\nu_2} + \frac{1}{\nu_1}\right)$. Some representative examples are shown in table 1.

To summarise, assuming (as in [3]) that the phase and flow diagram for the upper-half complex conductivity plane can be generated from an action of $\Gamma_0(2)$ which commutes with the RG, one deduces: (i) that all critical points are given by $\gamma(\sigma^*)$, where $\sigma^* = \frac{1+i}{2}$, with $\gamma \in \Gamma_0(2)$; (ii) the phase diagram of [4], [2] and [8]; (iii) the laws of corresponding states [4], [5]; and (iv) the selection rule $|p_1 q_2 - p_2 q_1| = 1$, dictating which transitions are allowed and which are forbidden. Lastly, the semi-circle law of [14] - [16] is compatible with, but not implied by, $\Gamma_0(2)$.

It should be noted that the full modular group does *not* provide the correct phase structure, since it would imply further critical points at the images of $\sigma = i$ and $\sigma = \frac{1+i\sqrt{3}}{2}$, under $\gamma \in \Gamma(1)$, which are not observed experimentally. The appearance of $\Gamma_0(2)$ is due to the extension of Kramers-Wannier duality $\sigma_{xx} \to 1/\sigma_{xx}$ to the whole complex plane. It was argued in [17] that this extension leads naturally to $\Gamma(1)$ acting on the upper-half complex plane, for a coupled clock model. This was applied to the quantum Hall effect in [18] and [19]. It appears to have been noted first in [1] that the subgroup $\Gamma_0(2)$ has the special property of preserving the parity of the denominator for rational $\nu = p/q$. The subgroup $\Gamma(2)$, consisting of all elements of $\Gamma(1)$ which are congruent to the identity, mod 2, was also considered in [2] and has been further investigated in [20]. Note however that there is no element of $\Gamma(2)$ which leaves $\sigma^* = \frac{1+i}{2}$ fixed, indeed there is no element of $\Gamma(2)$ which leaves *any* σ with $\infty > \Im(\sigma) > 0$ fixed.

It is a pleasure to thank Jan Pawlowski for discussions about the RG flow of the quantum Hall effect.

References

[1] C. A. Lütken and G. G. Ross, Phys. Rev. **B48**, 2500 (1993)

[2] C. A. Lütken, Nuc. Phys. **B396**, 670 (1993);
J. Phys. A : Math. Gen. **26**, L811-L817 (1993)

[3] C. P. Burgess and C. A. Lütken, Nuc. Phys. **B500**, 367 (1997)

[4] S. Kivelson, D-H. Lee and S-C. Zhang, Phys. Rev. **B46**, 2223 (1992)

[5] D-H. Lee, S. Kivelson and S-C. Zhang, Phys. Rev. Lett. **68**, 2386 (1992)

[6] R. A. Rankin, *Modular Forms and Functions*, C.U.P. (1977)

[7] D. E. Khmel'nitskii, Pis'ma Zh. Eksp. Teor. Fiz **38**, 454 (1983) (JETP Lett. **38**, 552 (1983)

[8] H. P. Wei, A.M. Chang, D. C. Tsui, A. M. M. Pruisken and M. Razeghi, Surf. Sci. **170**, 238 (1986)

[9] Y. Huo, R. E. Hetzel and R. N. Bhatt, Phys. Rev. Lett. **70**, 481 (1993)

[10] J. Zang and J.L. Birman, Phys. Rev. **B47**, 16305 (1993)

[11] R. Willett, J. P. Eisenstein, H. L. Störmer, D. C. Tsui, A. C. Gossard and J. H. English, Phys. Rev. Lett. **59**, 1776 (1987)

[12] R. B. Laughlin, Phys. Rev. Lett. **50**, 1395 (1983)
F. D. M. Haldane, Phys. Rev. Lett. **51**, 605 (1983)

[13] R. G. Clark, J. R. Mallett, S. R. Haynes, J. J. Harris and C. T. Foxon, Phys. Rev. Lett. **60**, 1747 (1988)

[14] A. M. Dykhne and I. M. Ruzin, Phys. Rev. **B50**, 2369 (1994)

[15] I. Ruzin and S. Feng, Phys. Rev. Lett. **74**, 154 (1995)

[16] I.M. Ruzin. N.R. Cooper and B.I. Halperin, Phys. Rev. **B53**, 1558 (1996)

[17] J. L. Cardy and E. Rabinovici, Nuc. Phys. **B205**, 1 (1982) J. L. Cardy, Nuc. Phys. **B205**, 17 (1982)

[18] A. Shapere and F. Wilczek, Nuc. Phys **B320**, 669 (1989)

[19] C. A. Lütken and G. G. Ross, Phys. Rev. **B45**, 11837 (1992)

[20] Y. Georgelin and J-C. Wallet, Phys. Lett. **A224**, 303 (1997); Y. Georgelin, T. Masson and J-C. Wallet, J. Phys. **A30**, 5065 (1997)

[21] D. Shahar, D. C. Tsui, M. Shayegan, E. Shimshoni and S. L. Sondhi, (cond-mat/9611011)

[22] D. Shahar, D. C. Tsui, M. Shayegan, R. N. Bhatt and J. E. Cunningham, Phys. Rev. Lett. **74**, 4511 (1995)
D. Shahar, D. C. Tsui, M. Shayegan, J. E. Cunningham, E. Shimshoni and S. L. Sondhi, (cond-mat/9607127) Solid State Comm. **102**, 817 (1997)

[23] M. Hilke, D. Shahar, S. H. Song, D. C. Tsui, Y. H. Xie and D. Monroe, (cond-mat/9708239)

[24] L. W. Wong, H. W. Jiang, N. Trivedi and E. Palm, Phys. Rev. **B51**, 18033 (1995)

[25] H. W. Jiang, R. L. Willett, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer and K. W. West, Phys. Rev. Lett. **65**, 633 (1990)

Table 1. Some examples of allowed transitions. The matrix γ maps the points $\sigma = 0$ and $\sigma = 1$ to the transition indicated in the leftmost column. Some representative experimental support (not exhaustive) is also indicated. The last two columns assume the semi-circle law (ρ is the resistivity).

Transition $\nu_1 \to \nu_2$	γ	Critical Conductivity	Critical Resistivity	σ at σ_{xx}^{Max}	ρ at ρ_{xx}^{Max}
$n+1 \to n$	$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$	$\frac{(2n+1)+i}{2}$	$\frac{(2n+1)+i}{2n^2+2n+1}$ (a)	$\frac{(2n+1)+i}{2}$	$\frac{(2n+1)+i}{2n(n+1)}$ (b)
$\frac{1}{2n+1} \to 0$	$\begin{pmatrix} 1 & 0 \\ 2n & 1 \end{pmatrix}$	$\frac{(2n+1)+i}{2(2n^2+2n+1)}$ (c)	$(2n+1)+i$ (d)	$\frac{1+i}{2(2n+1)}$	$(2n+1)+i\infty$
$\frac{n}{2n+1} \to \frac{n+1}{2n+3}$	$\begin{pmatrix} 1 & n \\ 2 & 2n+1 \end{pmatrix}$	$\frac{(4n^2+6n+3)+i}{2(4n^2+8n+5)}$	$\frac{(4n^2+6n+3)+i}{2n^2+2n+1}$	$\frac{(4n^2+6n+1)+i}{2(2n+1)(2n+3)}$	$\frac{(4n^2+6n+1)+i}{2n(n+1)}$ (e)
$\frac{3n+2}{4n+3} \to \frac{3n+5}{4n+7}$	$\begin{pmatrix} 3 & 3n+2 \\ 4 & 4n+3 \end{pmatrix}$	$\frac{(24n^2+58n+41)+i}{2(16n^2+40n+29)}$	$\frac{(24n^2+58n+41)+i}{18n^2+42n+29}$	$\frac{(24n^2+58n+29)+i}{2(4n+3)(4n+7)}$	$\frac{(24n^2+58n+29)+i}{2(3n+2)(3n+5)}$ (f)

(a) These points all lie on the semi-circle $\rho = i - \mathrm{e}^{i\theta} 0 \leq \theta \leq \pi$. For $n = 1$ see [21].
(b) Assumes $n \neq 0$.
(c) These points all lie on the semi-circle $\sigma = \frac{1}{2}(i - \mathrm{e}^{i\theta})$, $0 \leq \theta \leq \pi$.
(d) For $n = 0$ see [22] and [23], for $n = 1$ see [22] and [24], for $n = 2$ see [25].
(e) Assumes $n \neq 0$. For $n = 1, \ldots, 5$ and $n = -3, \ldots, -7$ see [14].
(f) For $n = 0$ see [14].

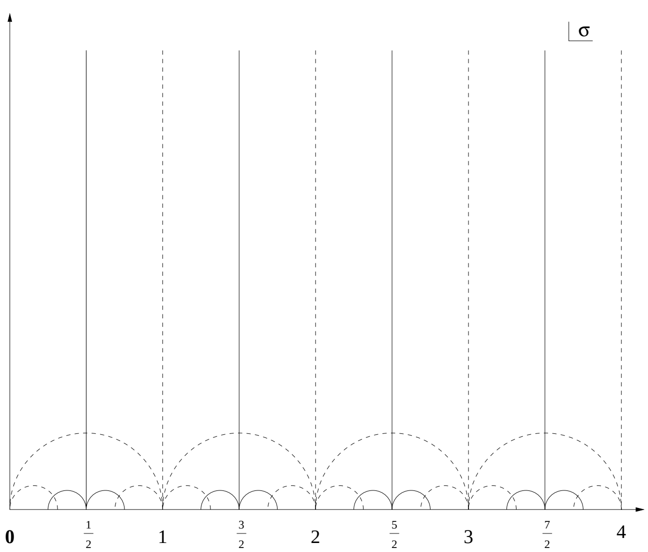

Fig. 1 The phase structure in the upper-half complex σ plane. The solid curves represent phase boundaries and the dotted curves allowed transitions. Points where dotted and solid lines cross are critical points.

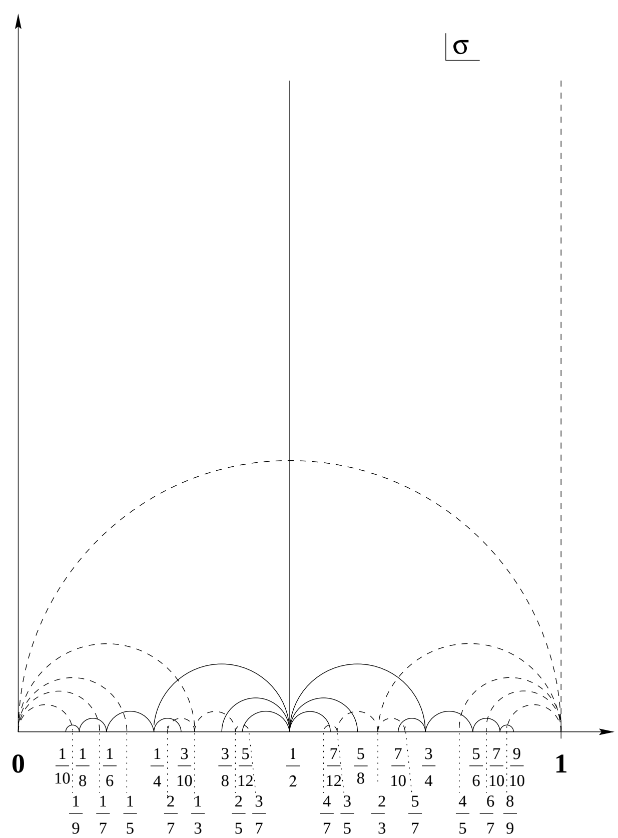

Fig. 2 A magnified view of the phase structure in the upper-half complex σ plane.